Biostatistics I – Introduction to Statistics and Experimental Design

Grace E. Kissling, Ph.D.
Biostatistics Branch

National Institute of Environmental Health Sciences

Outline

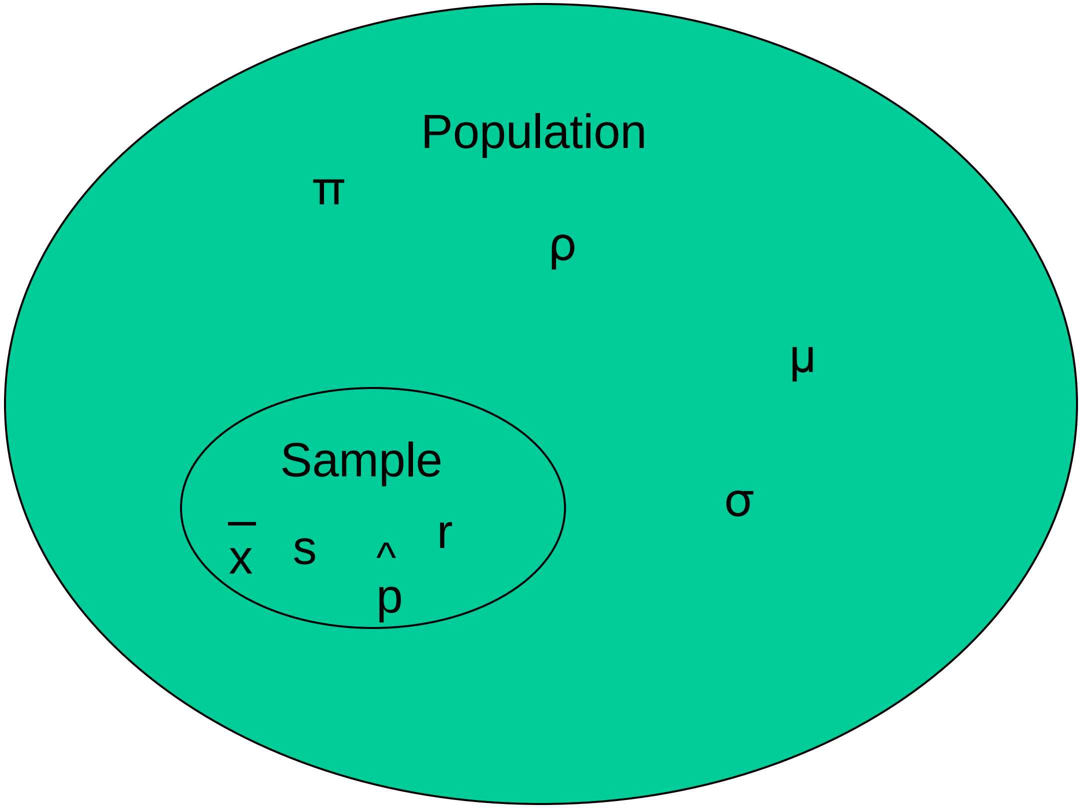
- Last time:
  - Study design
  - Levels of measurement
  - Numerical and graphical summaries
  - Sample size determination
- Today:
  - Estimation
  - Confidence intervals
  - Principles of hypothesis testing

Estimation

A <u>point estimate</u> gives a single value, such as a mean.

<u>Interval estimation</u> gives a range of likely values.

Population vs. Sample

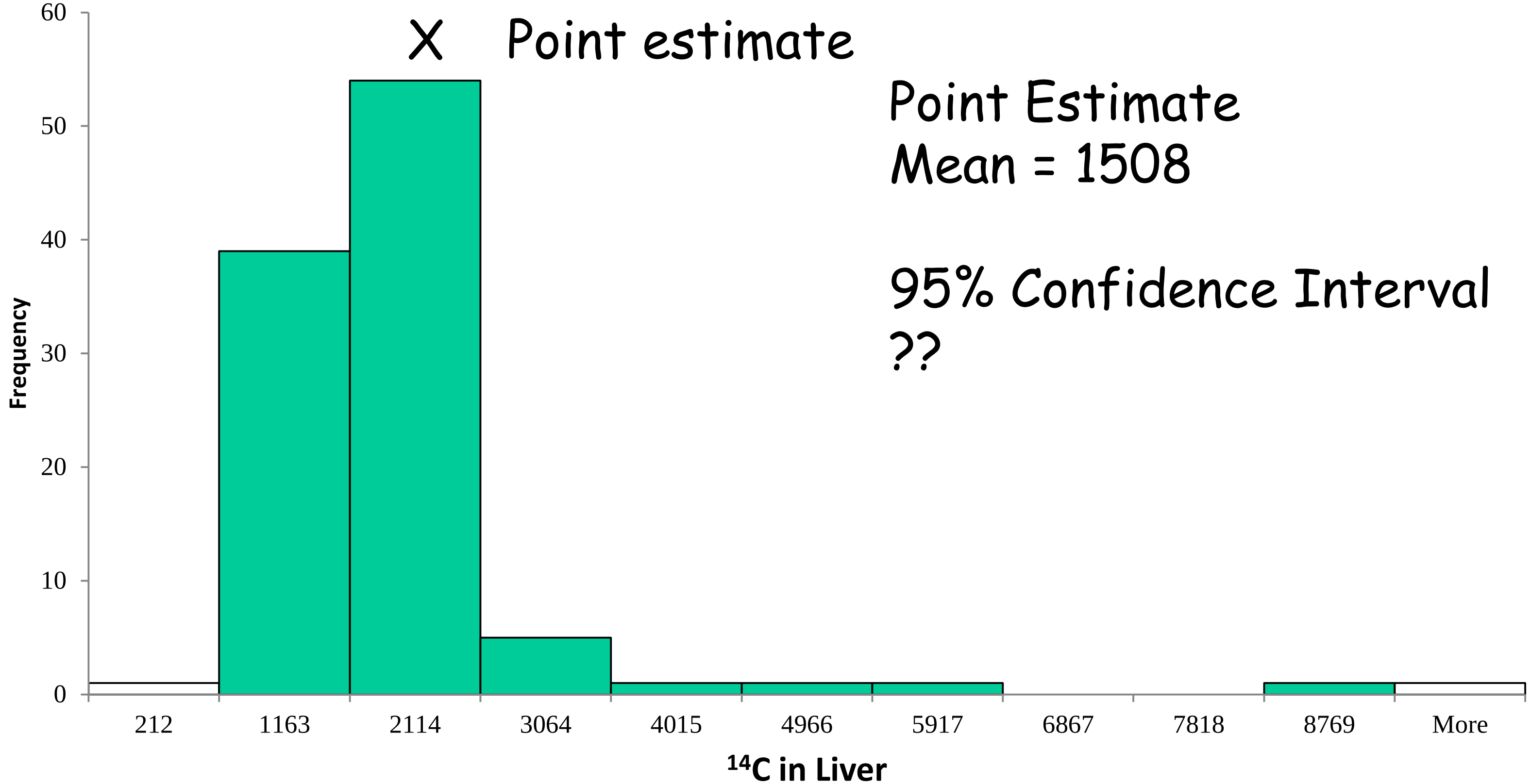


Confidence Intervals

A 95% confidence interval for the mean gives an interval that has 95% probability of capturing the mean of the population

..... meaning that if we were to conduct the experiment an infinite number of times, 95% of the time, the confidence interval that we construct will include the mean of the population.

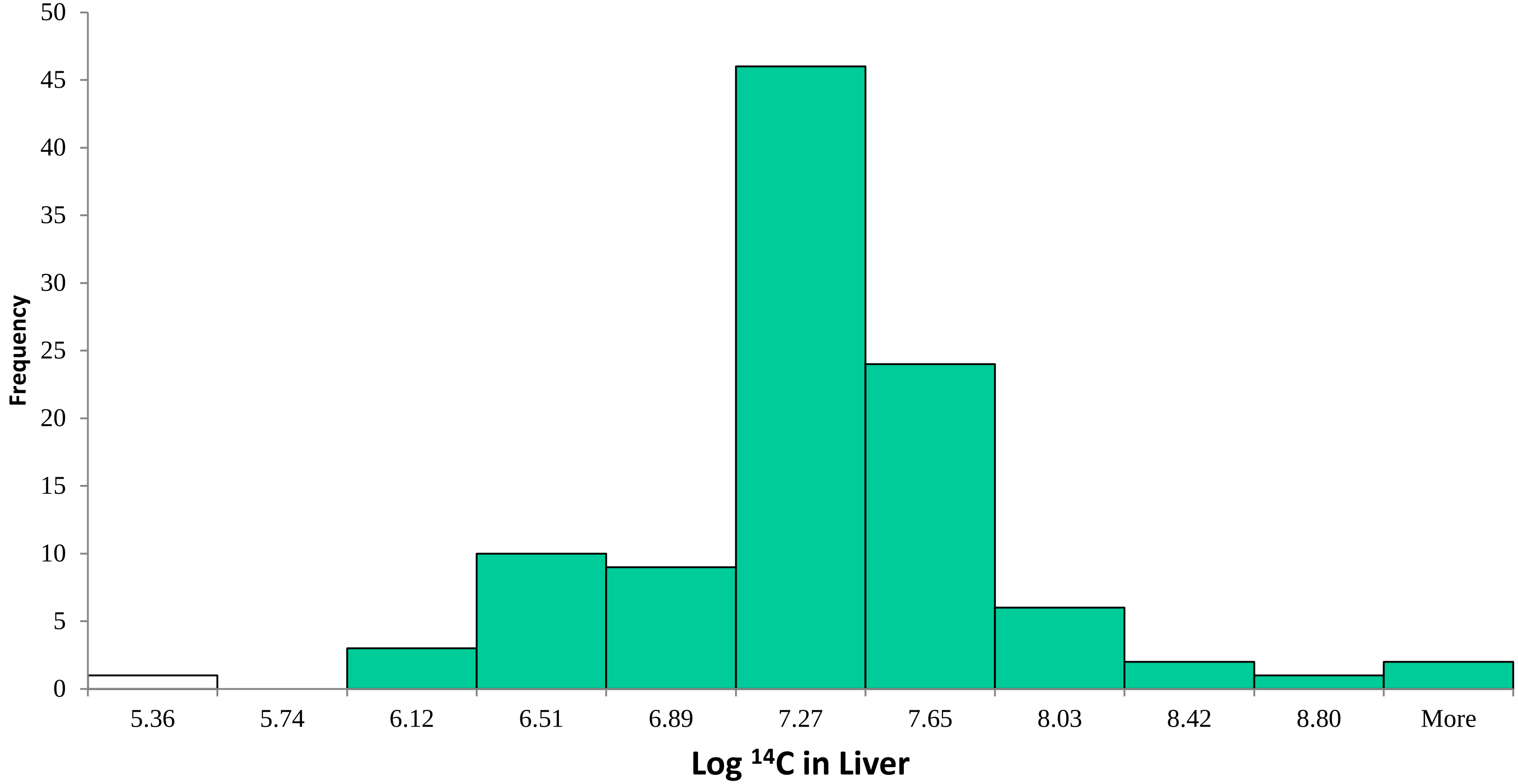
Example: Mean of Liver ^{14}C



Point Estimate
Mean = 1508

95% Confidence Interval
??

Example: Log(Liver ^{14}C)



Confidence Intervals (CI)

- For log(^{14}C in liver),

  Mean = 7.14, S.D. = 0.56, N = 104

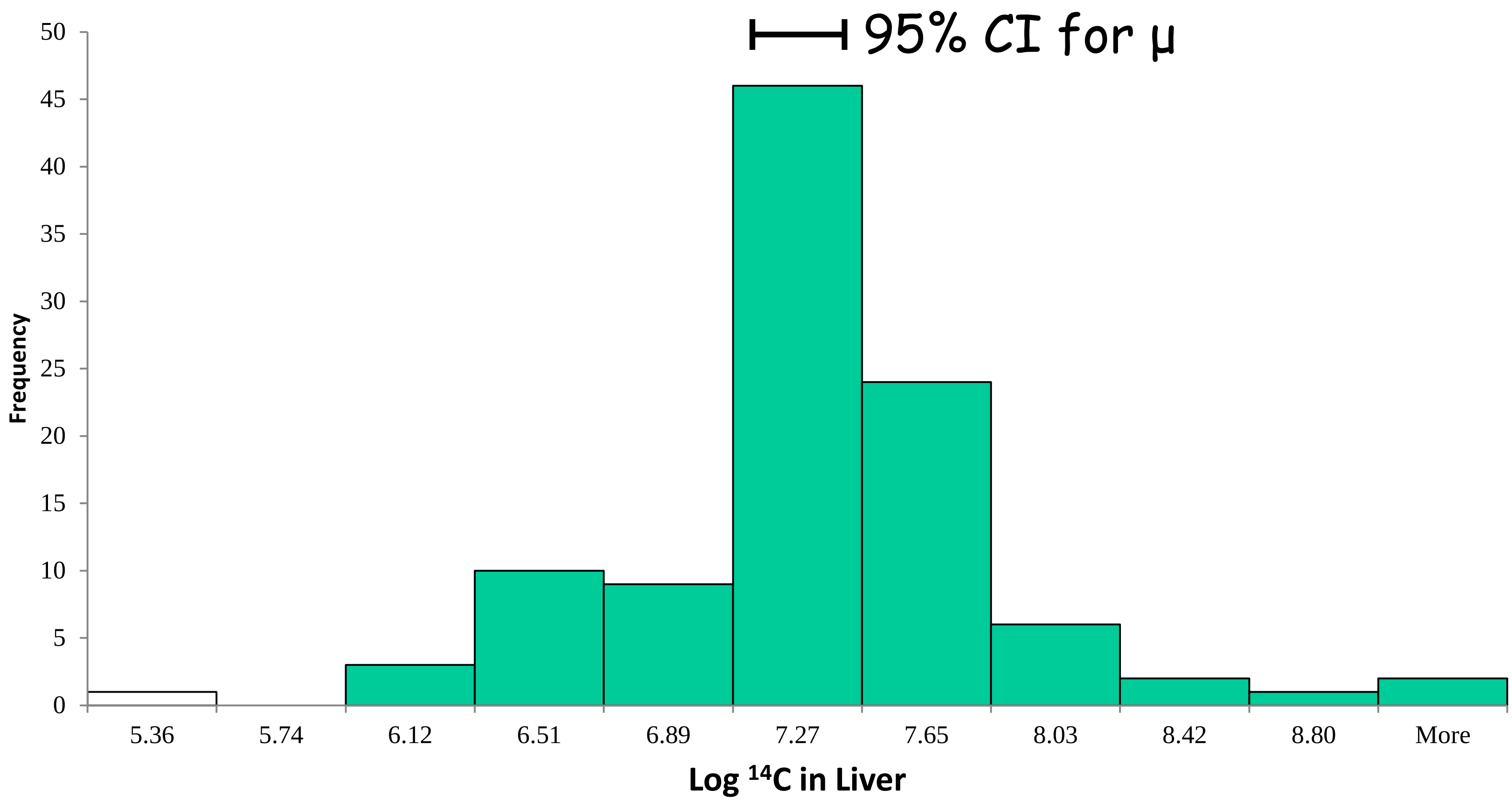
- The 95% CI for mean log(^{14}C in liver) is

$$\bar{x} \pm t_{103,.975} \times s.e.m. =$$

$$7.14 \pm 1.98 \times 0.56 / \sqrt{104}$$

$$(7.03, 7.25)$$

Example: Log(Liver ^{14}C)



Confidence Intervals

- For log(^{14}C in liver),
  Mean = 7.14, S.D. = 0.56, N = 104
- The 95% CI for mean log(^{14}C in liver) is

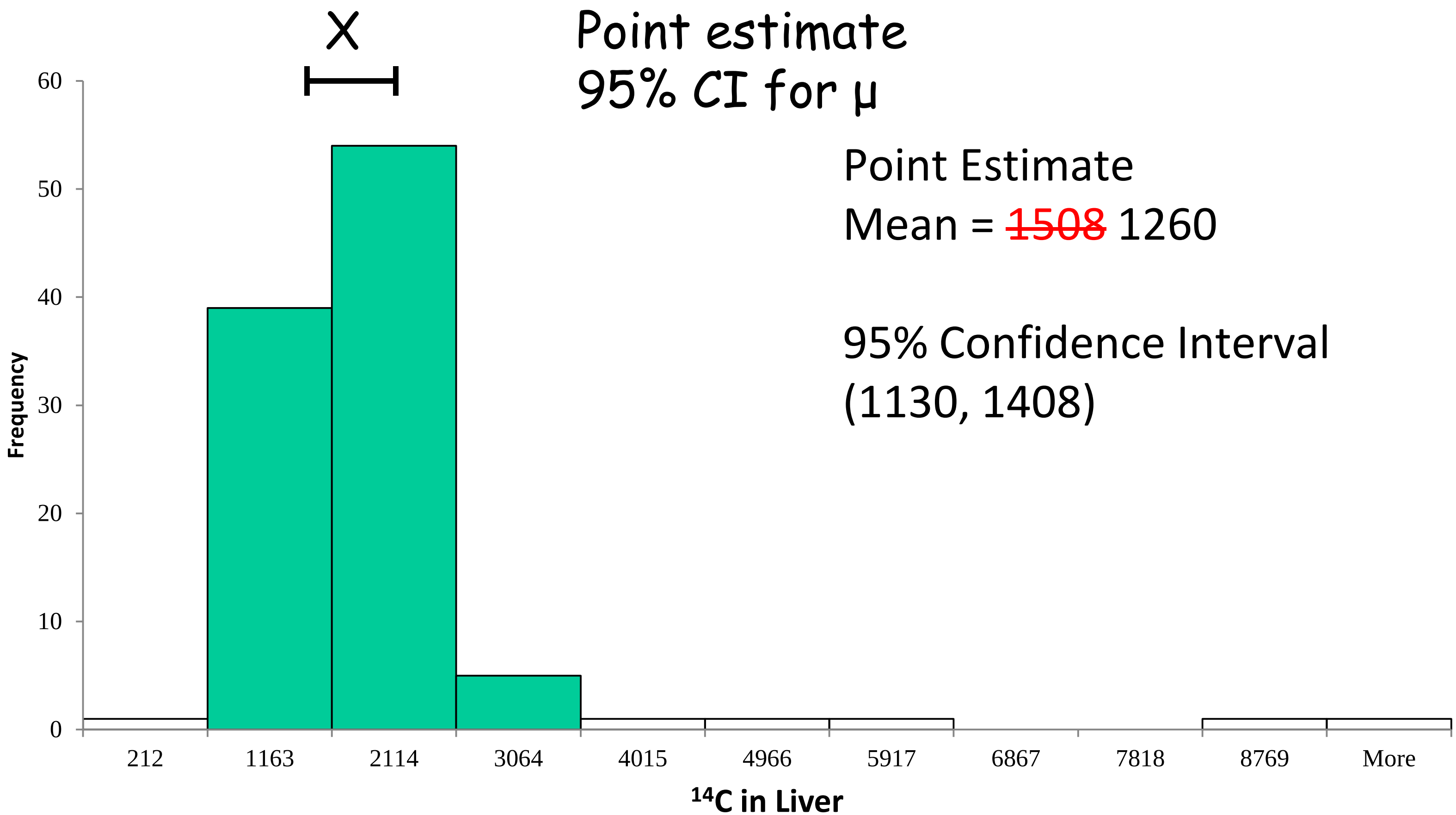
$$\bar{x} \pm t_{103,.975} \times s.e.m. =$$

$$7.14 \pm 1.98 \times 0.56 / \sqrt{104}$$

$$(7.03, 7.25)$$

Exponentiated, the 95% CI for mean ^{14}C in liver is (1130, 1408)

Example: Mean of Liver ^{14}C



Point estimate
95% CI for μ

Point Estimate
Mean = ~~1508~~ 1260

95% Confidence Interval
(1130, 1408)

Outliers

- Unusual values that require examination
- ***Do not*** automatically discard outliers
- Several methods exist for detecting outliers
  - Massey-Dixon test or Dixon's Q test
  - Grubbs' test
  - Exceeds 3 standard deviations from the mean
  - Exceeds 1.5 IQRs from 25^{th} or 75^{th} percentiles

Outliers

- Example: ^{14}C in liver (n = 7), Dixon's Q test
- 212, 403, 411, 433, 519, 524, 971

447

759

- Q = 447/759 = 0.59, look up in a Dixon's Q table to see if it exceeds the listed critical values
- For n = 7, the critical values are:
  - 0.507 at p = 0.10
  - 0.568 at p = 0.05
  - 0.680 at p = 0.01

0.59, outlier!

Outliers: Should I remove them?

- Legitimate reasons for removal:
  - Equipment malfunction
  - Impossible value
  - Error in data collection
  - Other mistake in the experiment

- Do not remove if:
  - An unusual value simply can't be explained
  - My data would "look better" if I removed it

Hypothesis Testing

- Null hypothesis
- Alternative hypothesis
- Test statistic
- P-value
- Conclusion

Hypotheses

- Null hypothesis, H_0
  - No difference, No effect, or No relationship
  - Always test H_0
    - assume H_0 true until there is sufficient evidence to the contrary

- Alternative hypothesis, H_1 or H_a
  - Usually, this is the research question

Test Statistic

- This is the evidence in favor of H_0
- Common test statistics have one of 4 well-known distributions:
  - Normal or z
  - Student's t
  - F
  - Chi-square

P-value

- P = Probability of the observed result or results more extreme, assuming H_0 is true

- One-sided or Two-sided P-value?
  - Can you predict *a priori* how groups will differ?
    - YES – use one-sided p
    - NO – use two-sided p

Conclusion

- If p is large, H_0 is supported.
- If p is small, H_0 is not supported, so we conclude that H_a is more likely correct.
- We typically use 0.05 to describe what is "large" ($p > 0.05$) and what is "small" ($p < 0.05$).

Keep in Mind....

- Our decision regarding H_0 is based on probabilities, so it could be incorrect.
- 0.05 is arbitrary.
- We use the significance level and the power to keep the probabilities of an incorrect decision low.

	H_0 is TRUE	H_0 is FALSE

	H_0 is TRUE	H_0 is FALSE
Reject H_0		
Accept H_0		

	H_0 is TRUE	H_0 is FALSE
Reject H_0	Type I error, False positive α	
Accept H_0		

	H_0 is TRUE	H_0 is FALSE
Reject H_0	Type I error, False positive α	
Accept H_0	Correct decision	

	H_0 is TRUE	H_0 is FALSE
Reject H_0	Type I error, False positive α	Correct decision $(1 - \beta)$, Power
Accept H_0	Correct decision	

	H_0 is TRUE	H_0 is FALSE
Reject H_0	Type I error, False positive α	Correct decision $(1 - \beta)$, Power
Accept H_0	Correct decision	Type II error, False negative β

Hypothesis Testing: Examples

Two-sample t-test

Paired t-test

Chi-square test

Fisher's exact test

Tests for normality

Hypothesis Testing: Mouse Body Weights

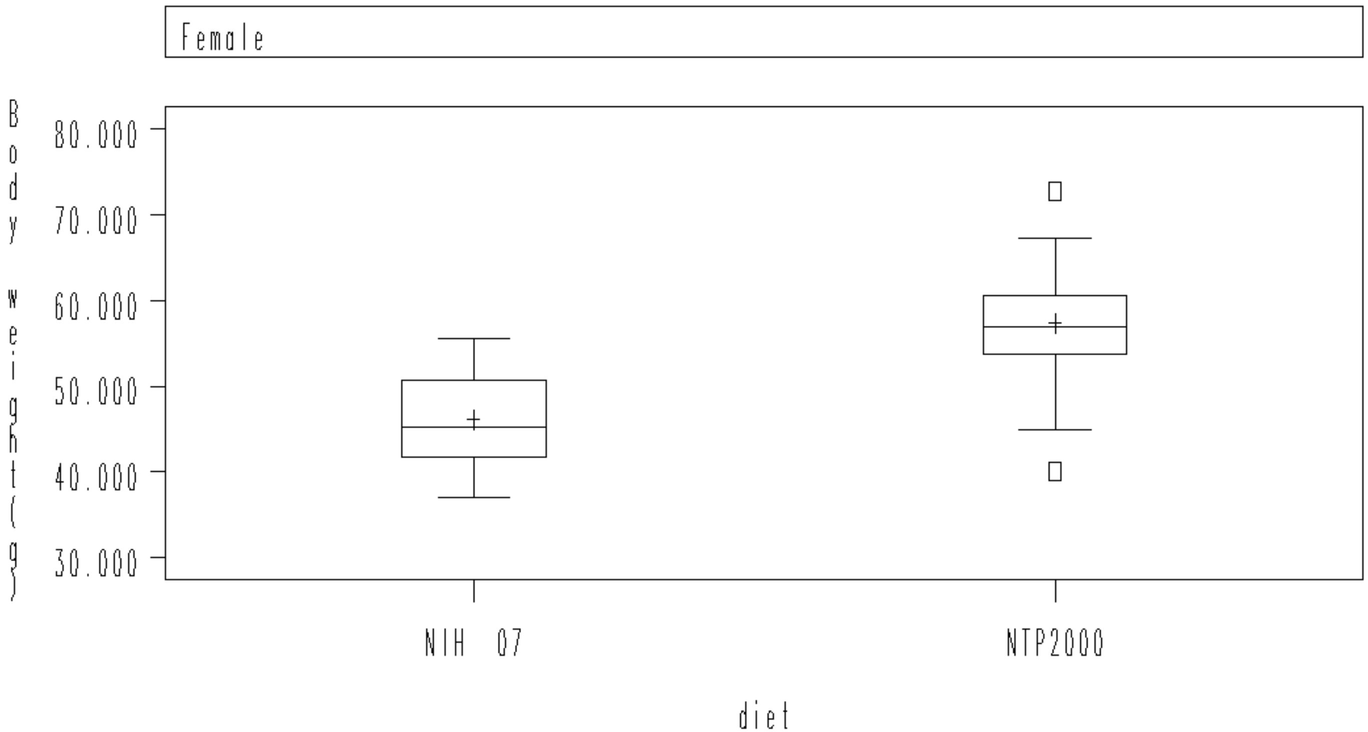
Body weights of 46 female mice on the NIH-07 diet and 49 female mice on the NTP 2000 diet were measured at one year of age.

Is there a difference in mean body weights between the two diet groups?

Two-sided p-value

Hypotheses

- H_0: Mean body weights are the same for the NIH-07 and NTP 2000 diet groups
- H_a: Mean body weights differ between the NIH-07 and NTP 2000 diet groups



N	46	49
Mean	46.12174	57.34082
Std Dev	5.410521	6.03462

Test Statistic and P-value

- NIH-07

  Mean = 46.1g, SD = 5.4g, N = 46
- NTP 2000

  Mean = 57.3g, SD = 6.0g, N = 49

Body weights are typically normally distributed

- Use a two-sample t-test

- $t(93) = 9.42$, $p < 0.0001$ (two-sided)

Conclusion

- If diet has no effect on one-year body weights, the probability of getting a mean difference of 11.2 g or more between the two diets is less than 0.0001.
- Because p is small ($p < 0.0001$), reject H_0 in favor of H_a
- Conclude that there is evidence that mean body weights of female mice at one year differ between the NIH-07 and NTP 2000 diet groups.

Selecting an Appropriate Test

The test statistic depends on:

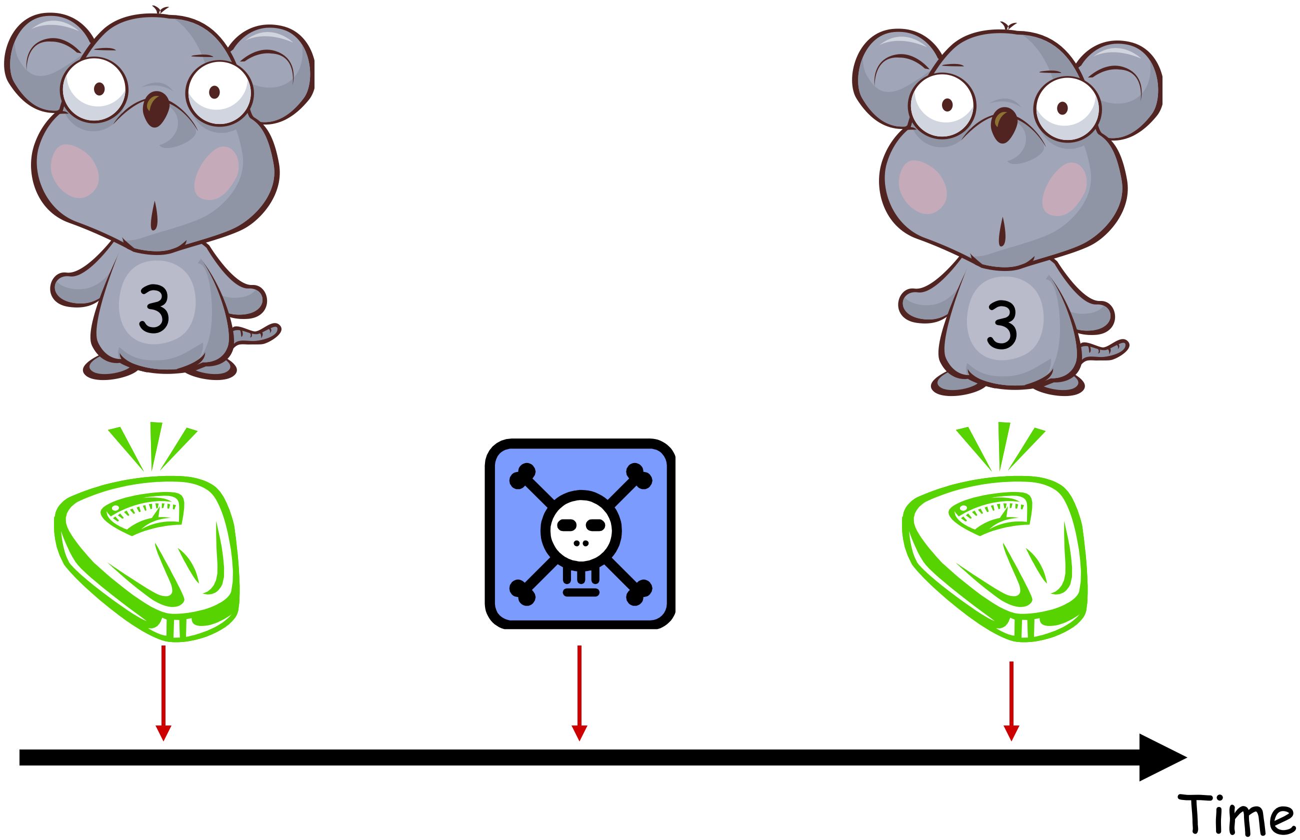
- Study design
- Hypotheses
- Level of measurement of data (nominal, ordinal, interval/ratio)
- Shape of the distribution

A New Study of Body Weights

- H_0: Mean body weights are not affected by treatment with Compound X
- H_a: Mean body weights are decreased after treatment with Compound X

One-sided p-value

Experimental Design



The Data

Mouse	Before X	After X	Difference, After - Before
1	36.3	35.0	-1.3
2	43.5	42.2	-1.3
3	32.0	32.6	0.6
4	50.4	50.6	0.2
5	52.1	51.5	0.6
6	56.3	54.2	-2.1
7	52.4	50.8	-1.6
Mean	46.1	45.3	-0.9
S.D.	9.1	8.7	1.0

Test Statistic and P-value

- Before: Mean = 46.1, SD = 9.1, N = 7
- After: Mean = 45.3, SD = 8.7, N = 7

Body weights are typically normally distributed

- Use a paired t-test

- $t(6) = 2.35$, $p = 0.029$ (one-sided)

Conclusion

- Assuming that Compound X has no effect on body weights, the probability of getting an average decrease of 0.9 g or more after administering Compound X is 0.029.
- Because p is small ($p = 0.029$), reject H_0 in favor of H_a
- Conclude that there is evidence that mean body weights of mice are lower after exposure to Compound X than they were before exposure.

What if I had ignored the study design?

- Before: Mean = 46.1, SD = 9.1, N = 7
- After: Mean = 45.3, SD = 8.7, N = 7

Body weights are typically normally distributed

- Use a two-sample t-test (ignores the pairing)

INCORRECT!!

- $t(12) = 0.18$, $p = 0.429$, Not Significant

Hypothesis Testing: Chi-square Test

H_0: Tumor rates are the same in Control and Treated animals
H_a: Tumor rates differ in Control and Treated animals

	Tumor	No Tumor	Total
Control	3	47	50
Treated	10	40	50
Total	13	87	100

$\chi^2 = 4.33$ with 1 degree of freedom (df)

P = 0.037

Reject H_0 because $0.037 < 0.05$. Conclude that there is a significant difference in tumor rates between Control and Treated animals.

Hypothesis Testing: Fisher's Exact Test

H_0: Tumor rates are the same in Control and Treated animals
H_a: Tumor rates are higher in Treated than Control animals

P = Probability of the observed data or data more extreme, if H_0 is true

	Tumor	No Tumor
Control	3	47
Treated	10	40

	Tumor	No Tumor
Control	2	48
Treated	11	39

	Tumor	No Tumor
Control	1	49
Treated	12	38

	Tumor	No Tumor
Control	0	50
Treated	13	37

Hypothesis Testing: Fisher's Exact Test

	Tumor	No Tumor
Control	3	47
Treated	10	40

Prob = 0.0283

	Tumor	No Tumor
Control	2	48
Treated	11	39

Prob = 0.0064

	Tumor	No Tumor
Control	1	49
Treated	12	38

Prob = 0.0009

	Tumor	No Tumor
Control	0	50
Treated	13	37

Prob = 0.00005

P-value = 0.0283 + 0.0064 + 0.0009 + 0.00005 = 0.0357

Hypothesis Testing: Fisher's Exact Test

$p = 0.0357$ (one-sided)

Because $p = 0.0357 < 0.05$, reject H_0 in favor of H_a that the tumor rate is higher in Treated animals than in Control animals.

Fisher's Exact Test and the Hypergeometric Distribution

	Tumor	No Tumor	
Control	3	47	50
Treated	10	40	50
	13	87	100

$$\text{Prob} = \frac{\binom{50}{3}\binom{50}{10}}{\binom{100}{13}} = 0.0283$$

Fisher's Exact Test and the Hypergeometric Distribution

	Tumor	No Tumor	
Control	2	48	50
Treated	11	39	50
	13	87	100

$$\text{Prob} = \frac{\binom{50}{2}\binom{50}{11}}{\binom{100}{13}} = 0.0064$$

Fisher's Exact Test and the Hypergeometric Distribution

	Tumor	No Tumor	
Control	1	49	50
Treated	12	38	50
	13	87	100

$$\text{Prob} = \frac{\binom{50}{1}\binom{50}{12}}{\binom{100}{13}} = 0.0009$$

Fisher's Exact Test and the Hypergeometric Distribution

	Tumor	No Tumor	
Control	0	50	50
Treated	13	37	50
	13	87	100

$$\text{Prob} = \frac{\binom{50}{0}\binom{50}{13}}{\binom{100}{13}} = 0.00005$$

Fisher's Exact Test and the Hypergeometric Distribution

P-value = 0.0283 + 0.0064 + 0.0009 + 0.00005 = 0.0357

Hypothesis Testing: Test for Normality

Are my data normally distributed?

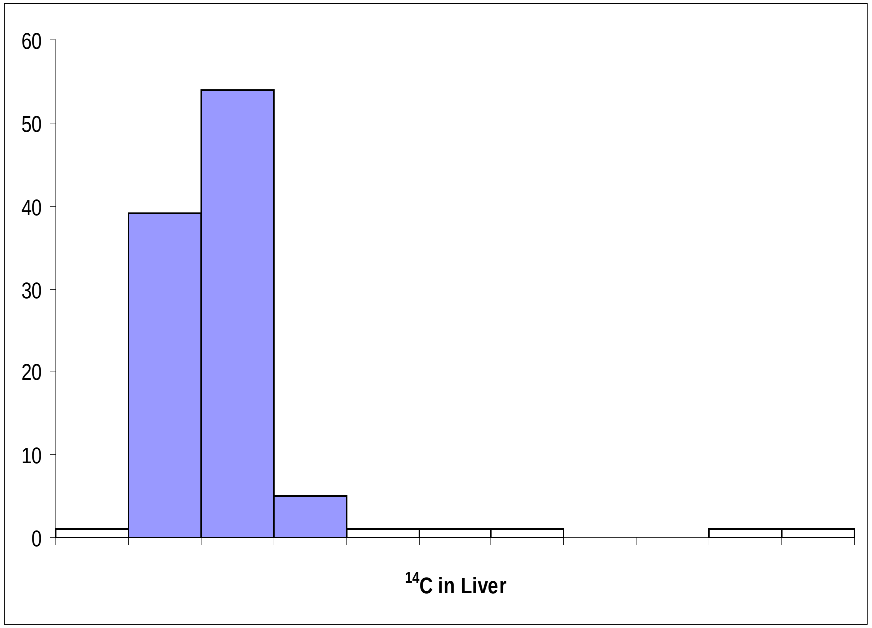
H_0: The data are normally distributed.
H_a: The data are not normally distributed.

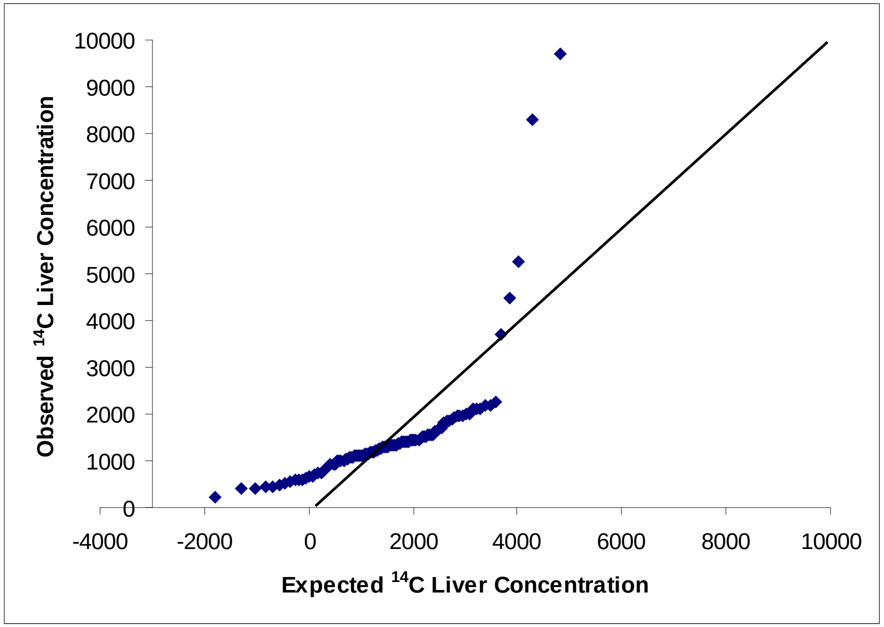
There are many tests for normality

- Shapiro-Wilks test
- Kolmogorov-Smirnov test
- Liliefors test
- Cramer-von Mises test
- Anderson-Darling test

Computation is tedious, so we rely on software to do the work.

Test for Normality: Recall ^{14}C in Liver





Hypothesis Testing: Test for Normality

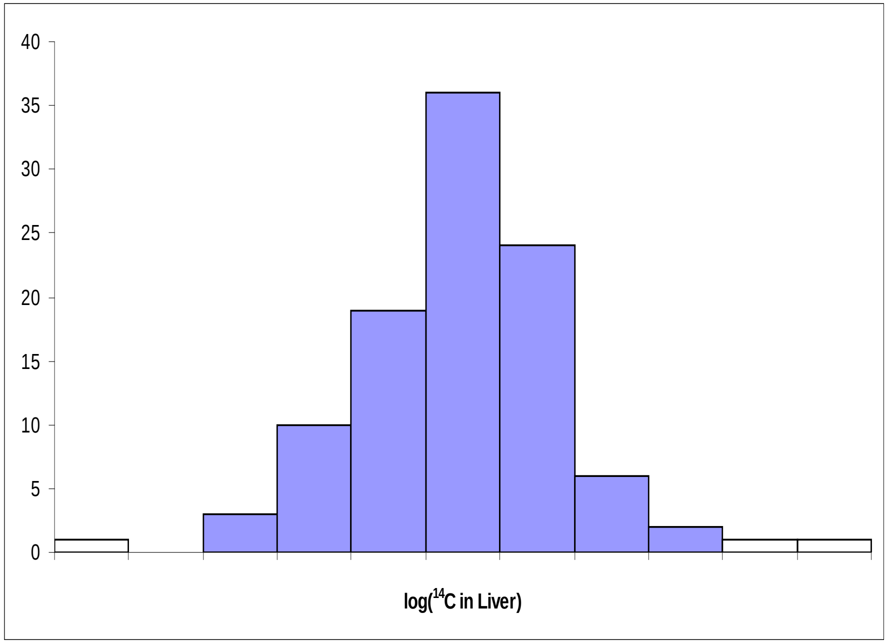
Are my data normally distributed?

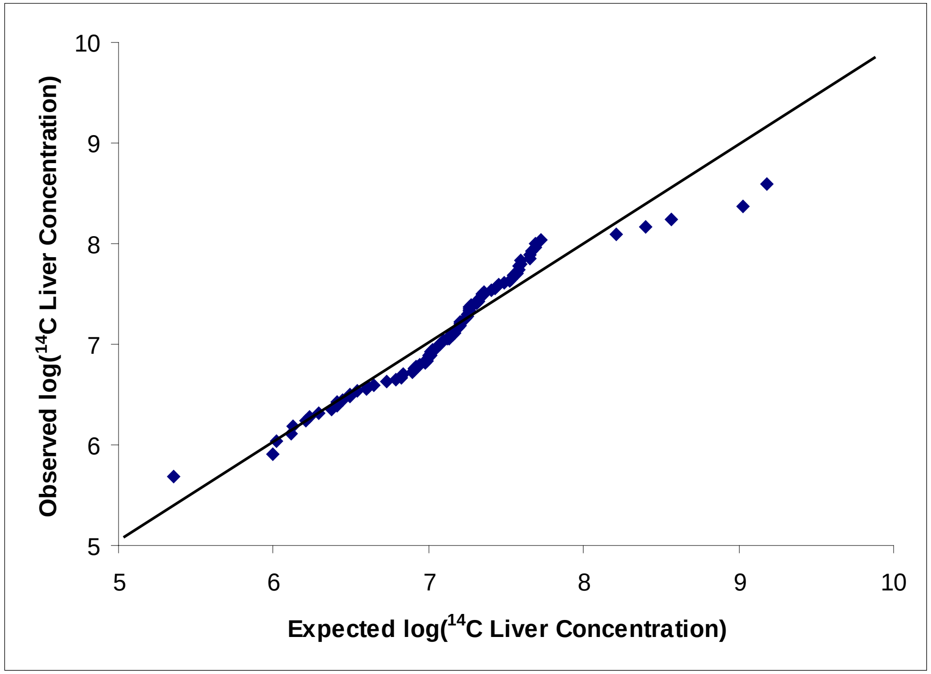
H_0: The data are normally distributed.
H_a: The data are not normally distributed.

^{14}C in liver:
Shapiro-Wilks statistic = 0.54 (using software)
$P < 0.0001$

Reject H_0: the data are not normally distributed.

Test for Normality: Recall log(^{14}C in Liver)





Hypothesis Testing: Test for Normality

Are my data normally distributed?

H_0: The data are normally distributed.
H_a: The data are not normally distributed.

Log(^{14}C in liver):
Shapiro-Wilks statistic = 0.98
P = 0.44

Accept H_0: the data are normally distributed.

End of Biostatistics I

While I have presented some methods for analyzing data, one must be careful in applying them.

In Biostatistics II, we will take a look at some of the caveats.

"IF you torture the data long enough, it will confess. But there is no guarantee that it will tell you the truth."

--from Berk, Regression Analysis: A Constructive Critique, 2004.